

CLAIMS

What is claimed is:

1. A system comprising:
 - at least one terminal node;
 - at least one network resource, said network resource having at least one redundant matching resource;
 - a computer that transfers tasks from said network resource to said redundant matching resource if said network resource fails, and that balances loads between said network resource and said redundant matching resource; and
 - a communication medium connecting said computer, said terminal node, said network resource and said redundant matching resource, said communication medium having at least one redundant communication path between said terminal node, said network resource and said redundant matching resource.
2. The system of claim 1, wherein said communication medium is selected from a group consisting of a local area network (LAN), a wide area network (WAN), an Ethernet based network, an Internet protocol (IP) based network, an asynchronous transfer mode (ATM) based network, a peripheral component interconnect (PCI) based network, and an InfiniBand based network.
3. The system of claim 1, wherein said communication medium further comprises at least one communication element.

4. The system of claim 3, wherein said at least one communication element is selected from a group consisting of a network switch and a cache control node.

5. The system of claim 1, wherein said at least one network resource is selected from a group consisting of a storage device, a redundant array of independent disks (RAID), a cache, a file system, and a location independent file system.

6. The system of claim 1, wherein said at least one redundant matching resource is selected from a group consisting of a storage device, a redundant array of independent disks (RAID), a file system, and a location independent file system.

7. The system of claim 1, wherein said at least one network resource and said at least one redundant matching resource are continuously interchangeable.

8. The system of claim 1, wherein said computer load balances by concurrently assigning tasks to said at least one network resource and said at least one redundant matching resource.

9. A system comprising:

a plurality of terminal nodes;

a plurality of network resources;

a plurality of redundant resources, each of said plurality of network resources closely matching at least one of said plurality of redundant resources;

a computer that moves tasks from a failed network resource to a redundant resource that closely matches the failed network resources, and that balances loads between said plurality of network resources and said plurality of redundant resources; and

a communication medium connecting said computer, said terminal nodes, said network resources and said redundant resources, said communication medium having at least one redundant communication path between said terminal nodes, said network resources and said redundant resources.

10. The system of claim 9, wherein said computer, after receiving a request for using said network resources, balances the load by:

assigning at least one network resource from said plurality of network resources to the request;

determining a communication path in said communication medium; and

informing the requestor of the assigned network resource and the determined communication path.

11. The system of claim 10, wherein the assigning of at least one network resource comprises:

determining the least loaded network resource of the requested resource type;

if the number of available network resources is larger than one, executing a selection function to determine the at least one network resource to be assigned, and assigning that resource to the requestor; and

if the number of available network resources is equal to one, assigning the available resource to the requestor.

12. The system of claim 11, wherein the selection function is selected from a group comprising a round robin function, a weighted round robin function, a random function, a least loaded function and a least recently used function.

13. The system of claim 10, wherein determining a communication path in said communication medium comprises:

determining the least loaded communication path to said assigned network resource;

if the number of available communication paths is larger than one, executing a selection function to determine a communication path to be used, and informing requestor of selected communication path; and

if the number of available communication paths is equal to one, informing requestor of available communication path.

14. The system of claim 13, wherein the selection function is selected from a group comprising a round robin function, a weighted round robin function, a random function, a least loaded function and a least recently used function.

15. The system of claim 9, wherein, if said computer receives a failure notification, said computer attempts to re-balance loads by:

determining the type of failure that caused the failure notification;

if the computer determines that a communication path in said communication medium has failed, then:

if no alternative communication path is available, the computer issues an error notification;

otherwise, the failed communication path is eliminated from further use, and the load is redistributed;

if the computer determines that a network resource has failed, then:

if no alternative network resource is available, the computer issues an error notification;

otherwise, the failed network resource is eliminated from further use, and the load is redistributed.

16. A method for balancing loads in a network system containing a plurality of communication paths, a plurality of redundant communication paths, a plurality of network resources of differing types each of said network resources

having at least one closely matching network resource that may be used as a redundant network resource, the method comprising:

receiving a request for access to one of said plurality of network resources;

assigning at least one network resource from said plurality of network resources to the request;

assigning one of said plurality of communication paths to the request; and

informing the requestor of the assigned network resource and the assigned communication path.

17. The method of claim 16, wherein the assigning of at least one network resource comprises:

determining the least loaded network resource of the requested resource type;

if the number of available network resources is larger than one, executing a selection function to determine the at least one network resource to be assigned, and assigning that network resource to the requestor; and

if the number of available network resources is equal to one, assigning the available network resource to the requestor.

18. The method of claim 17, wherein said selection function is selected from a group comprising a round robin function, a weighted round robin

function, a random function, a least loaded function and a least recently used function.

19. The method of claim 16, wherein assigning a communication path comprises:

determining the least loaded communication path to said assigned network resource;

if the number of available communication paths is larger than one, executing a selection function to determine a communication path to be used, and informing requestor of selected communication path; and

if the number of available communication paths is equal to one, informing requestor of available communication path.

20. The method of claim 19, wherein said selection function is selected from a group comprising a round robin function, a weighted round robin function, a random function, a least loaded function and a least recently used function.

21. A method for re-balancing loads in a network system containing a plurality of communication paths, a plurality of redundant communication paths, a plurality of network resources of differing types, each of said network resources having at least one closely matching network resource which may be used as a redundant network resource, said method comprising:

determining the type of failure that caused the failure notification;

if a communication path has failed, then:

if no alternative communication path is available, issuing an error notification;

otherwise, the failed communication path is eliminated from further use, and the load is redistributed;

if a network resource has failed, then:

if no alternative network resource is available, issuing an error notification;

otherwise, the failed network resource is eliminated from further use, and the load is redistributed.

22. A computer software product for balancing loads in a network system containing a plurality of communication paths, a plurality of redundant communication paths, a plurality of network resources of differing types, each of said network resources having at least one closely matching network resource which may be used as a redundant network resource, the computer program product comprising:

software instructions for enabling the network system to perform predetermined operations, and a computer readable medium bearing the software instructions, said predetermined operations comprising:

receiving a request for access to one of said plurality of network resources;

assigning at least one network resource from said plurality of network resources to the request;

assigning one of said plurality of communication paths to the request; and

informing the requestor of the assigned network resource and the assigned communication path.

23. The computer software product of claim 22, wherein assigning of at least one network resource comprises:

determining the least loaded network resource of the requested resource type;

if the number of available network resources is larger than one, executing a selection function to determine the at least one network resource to be assigned, and assigning that network resource to the requestor; and

if the number of available network resources is equal to one, assigning the available network resource to the requestor.

24. The computer software product of claim 23, wherein said selection function is selected from a group comprising a round robin function, a weighted round robin function, a random function, a least loaded function and a least recently used function.

25. The computer software product of claim 22, wherein assigning a communication path comprises:

determining the least loaded communication path to said assigned network resource;

if the number of available communication paths is larger than one, executing a selection function to determine a communication path to be used, and informing requestor of selected communication path; and

if the number of available communication paths is equal to one, informing requestor of available communication path.

26. The computer software product of claim 25, wherein said selection function is selected from a group comprising a round robin function, a weighted round robin function, a random function, a least loaded function and a least recently used function.

27. A computer software product for re-balancing loads in a network system containing a plurality of communication paths, a plurality of redundant communication paths, a plurality of network resources of differing types, each of said network resources having at least one closely matching network resource which may be used as a redundant network resource, the computer program product comprising:

software instructions for enabling the network system to perform predetermined operations, and a computer readable medium bearing the software instructions, the predetermined operations comprising:

determining the type of failure that caused the failure notification;

if a communication path has failed, then:

if no alternative communication path is available, issuing an error notification;

otherwise, the failed communication path is eliminated from further use, and the load is redistributed;

if a network resource has failed, then:

if no alternative network resource is available, issuing an error notification;

otherwise, the failed network resource is eliminated from further use, and the load is redistributed.

28. A redundant network system capable of using redundant elements for the purpose of load balancing the system comprising:

at least one client node;

at least two network switches providing alternate connection paths to said client node;

at least two cache control nodes capable of supporting an address resolution protocol and capable of load balancing storage control nodes, said cache control nodes connected to said network switches; and

at least two storage control nodes, said storage control nodes connected to at least said network switches.

29. The system of claim 28, wherein said address resolution protocol is executed on one of said cache control nodes.

30. The system of claim 28, wherein one of said cache control nodes is used as a redundant cache control node.

31. The system of claim 30, wherein the second cache control node receives address resolution protocol information from the first cache control node.

32. The system of claim 31, wherein, when one of said cache control nodes fails, the remaining cache control node executes said address resolution protocol.

33. The system of claim 28, wherein one of said cache control nodes generates a media access control address corresponding to a storage control node.

34. The system of claim 33, wherein said media access control address is for the storage control node which is the next to receive the access request based on a load balance function executed by one of the cache control nodes.

35. The system of claim 34, wherein said load balancing function is selected from a group comprising a round robin function, a weighted round robin function, a random function, a least loaded function and a least recently used function.

36. The system of claim 34, wherein said media access control address is further based on the specific network path to be used for said client node to connect to said storage control node.

37. The system of claim 33, wherein said media access control address is for the network path which is the next to receive the access request based on a load balance function executed by one of the cache control nodes.

38. The system of claim 37, wherein said load balancing function is selected from a group comprising a round robin function, a weighted round robin function, a random function, a least loaded function and a least recently used function.